ESG Equity Extensions

Mathematical Finance Company [1]

March 21, 2003

Contents

Chapter 1

Simulation

1.1 Introduction

In simulation, we start out by thinking we have the true model of nature or the economy or the economic process and we know everything. We don't have any observation problems, we know the parameters, we know the models, we know how to simulate.

We have as our theoretical model the idea that there are hidden variables which we can model as coming from such joint distribution like the multivariate normal. We use the latter because it has nice properties especially for modeling many random variables at once.

We have a model though that individual data series that we see in the economy, like the S & P 500 or the price of Ford stock are not normally distributed nor are simple transforms of them like the difference in natural logarithms of them at adjacent time points. Nor is the arithmetic return from one month end to the next month end normally distributed.

So we have a theoretical model that we can simulate with which starts out with the hidden multivariate normal distribution of many variables and simulates those first. Then for each such variable we transform it so as to produce the observable variables such as Ford stock price or the return of Ford stock price which we also classify as an observable because the transformation involves no estimated parameters.

So step one is generate all the normal variables at once for a given time period using a multivariate normal distribution. This models their correlation. Now we take each individual normal and turn it into an observable. The way we do this is by matching the univariate normal distribution's cumulative distribution to the

univariate cumulative distribution of some other distribution like logistic. At this point, we know the parameters of the logistic or final distribution as we might call it or final output distribution.

We also know the mean and variance of the inner core normal. In this model, it doesn't actually matter what those are as long as the variance isn't zero. The reason is that the parameters of the "output" univariate distribution will undo any effect of the mean or non-zero variance of the normal.

So we say that the cumulative normal (of the univariate normal distribution) up to the simulated value x equals the cumulative distribution of the "output" distribution, say the logistic, whose value we call y. The distribution of y values doesn't depend on the mean and variance of the normal as long as we use the same mean and variance for x's cumulative distribution to transform x to y as we used to simulate x.

We thus solve for y by equating its logistic or whatever cumulative density to the cumulative density of x, with the same mean and variance for the cumulative of x as are used to generate x. And as long as that is done and x is univariate normal, it doesn't matter what mean and variance were used to generate x as long as the variance wasn't zero.

So we solve for y and that is now our output variable. The variable y might however be something now like the change in the logarithm of the stock price or index over some period of time, like a month.

Note that the transformation has to pick a specific time interval for it to be based on which has some specified logistic. It is possible to define a continuous in time output variable but it will only be logistic with specified parameters over some specific interval.

1.2 MVN

Let x be an n by 1 column vector. The 1 by n row vector x' is given by

$$x' = (x_1, ..., x_n) \tag{1.1}$$

We assume that x is multivariate normal, MVN, with probability density function

$$f(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \tag{1.2}$$

1.3 Univariate Distributions

For each x_i we transform to a variable y_i, defined by the following procedure. Let μ_i and σ_i be the parameters of the univariate distribution for x_i. Let $F_i(x_i; \mu, \sigma^2)$ be the normal cumulative distribution with mean μ and variance σ^2 for variable x_i. We have

$$F_i(x_i; \mu, \sigma^2) = \int_{-\infty}^{x_i} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu_i)^2/\sigma^2)} \tag{1.3}$$

Let $G_i(y_i; \eta)$ be the cumulative distribution function of a single variable with parameter column vector η of dimension k_i by 1. Let the corresponding probability density function be $g(y_i; \eta)$, with appropriate modifications made for point masses, such as including a δ function in g. Let y_i be determined by the equation

$$F_i(x_i; \mu, \sigma^2) = G_i(y_i; \eta) \tag{1.4}$$

We solve the above equation for y_i conditional on the parameter vector η after simulating the vector x including the value x_i. In this approach, G_i can be any univariate distribution

1.4 Discrete time simulation

We can simulate in discrete time, most simply when the time intervals are equal.

1.5 Combining with other variables

We can combine the variables generated in this manner with other variables, which themselves may follow a continuous time process. The vector x of MVN for transforming to logistics can be generated as part of that process, and we can for example choose to make that subvector have a mean of zero and standard deviations of one over the time intervals at which we wish results. [1] This sub-vector then furnishes a vector of MVN for the above transformation. For the case of fixed time intervals for the desired logistics, we simply specify the corresponding logistic distribution parameters for an interval of that size and use the generated MVN over that period.

[1] This is not really necessary.

Chapter 2

Esimation of Parameters

For estimating parameters we have an inverse direction from simulation. In the case of simulation we start first with the simulation of the multivariate normal vector of variables first. Then we transform those to the individual output variables. This can be done by a transformation using cumulative density functions of the normal and any univariate distribution, such as a logistic.

In the case of estimation, we start with data that is already transformed so to speak by "nature" or the economy. Our job as experimentalists or as econometricians is then to reverse engineer nature. So we get from nature or the economy, variables that individually we identify as logistic or whatever.

We believe there is an inner core of correlation that is hidden by the logistic or univariate distributions of individual data series, such as the S & P 500, Ford stock price, etc.

We can't observe the correlation matrix directly nor estimate it directly from the observed univariate series, e.g. Ford stock price at month end, S & P 500 at month end from January 1950 to January 2003, or whatever period we use or data series we observe.

Because these invididual data series are logistic or whatever, we can't directly estimate the correlation matrix of them. The multivariate logistic is a very constrained distribution that we don't want to be restricted by. We might also want to use other univariate distributions than logistic.

So to uncover the hidden correlation of this collection of individual series, modeled by a collection of independent or seemingly independent univariate distributions, we want to transform each individual data series so that the transformed series is univariate normal. We then make the leap of faith, or assumption that these transformed series are in fact multivariate normal, MVN, and that we can estimate their correlation matrix by standard means for a MVN given a collection

of individual normals that are further assumed to be multivariate normal. As is well known, that is not guaranteed, but as modelers of nature or the economy we make that our assumption because its easier to build a model. If that doesn't work we can try something else.

So as econometricians our first problem is to estimate univariate distributions for each series that comes to us from nature or the economy, the SPX, Ford, whatever. We in fact first transform this by using differences in the natural logarithms of the series or constructing the arithmetic returns. Because these transforms involve no parameters to be estimated we think of this as still being the data that comes to us from nature or the economy.

2.1 Estimation of Parameters

Suppose we are given a vector of data y_t, t=1,...,T, where y_t is a vector of dimension n by 1. For each i, we take the time series y_{ti}, t=1,...,T, where y_{ti} is the i-th component of y_t. Given this data we can estimate a univariate distribution separately for each i. This can be done using standard univariate methods. Let the estimated parameters for each distribution be given by the vector η_i, where η_i is a k_i by 1 column vector.

Given these parameters, we can now transform to a distribution that is univariate normal. We discuss two ways to do this. One is to specify that the univariate normal has mean zero and variance 1. We then solve for x_{it} for t=1,...,T such that

$$\int_{-\infty}^{x_{it}} dz \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(z)^2} = G_i(y_{it}; \eta_i) \tag{2.1}$$

This gives us a set of vectors x_t, t=1,...,T. We now estimate the correlation matrix among these using standard methods. Let this correlation matrix of the vector x_t at any time point be denoted W.

We simulate with this approach a MVN with mean zero, and unit standard deviations with correlation matrix W. We then transform each y_i individually using $G_i(.,b_i)$.

As an alternative, we could specify a vector μ for the MVN and a vector of standard deviations σ. These could be the values estimated in the sample.

As a short cut to estimate the correlation matrix, we can use the correlation matrix of the raw data. When dealing with some data, we may wish to take a transformation like the natural logarithm first. If we deal with stock data, for example we can first take the change in the natural logs of the stock prices and treat those as the y's. In this form, we may use the correlation matrix of the y's as a quick estimate of the correlation matrix of the normal distribution.

Chapter 3

Statistics

3.1 Univariate Logistic Distribution

For any y between $-\infty$ and ∞, let the cumulative distribution of y given parameters α and β be

$$G(y;\alpha,\beta) = \frac{1}{1+e^{-(x-\alpha)/\beta}} \tag{3.1}$$

This can also be written

$$G(y;\alpha,\beta) = \frac{1}{2}(1 + tanh(\frac{1}{2}(x-\alpha)/\beta)) \tag{3.2}$$

G is the cumulative logistic distribution and y is said to be logistically distributed or to have the logistic distribution.

3.2 Logistic Normal Conversion

To do conversion we equate the cumulative distribution function of the Normal to that of the Logistic.

$$\int_{-\infty}^{x_{it}} dz \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(z)^2} = \frac{1}{1+e^{-(y_{it}-\alpha_i)/\beta_i}} \tag{3.3}$$

If we are simulating we simulate x with mean 0 and variance 1 and then use our model assumptions for α_i and β_i for series i, e.g. Ford or S & P 500 or whatever.

If we are estimating, we first estimate α_i and β_i from a time series of y_{it} and then we convert to x using the above formula given the observed value of y_{it}.

In both cases y_{it} is the return, log or arithmetic return, not the stock price or index value itself. For example x and y are both varying over $-\infty$ to ∞.

The above normal has zero mean and variance 1 which may seem counter-intuitive. Why doesn't it have a mean and variance corresponding to that of the logistic?

The reason is it wouldn't matter. Suppose we use some μ and σ in the transformation

$$\int_{-\infty}^{x_{it}} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu)^2/\sigma^2} = \frac{1}{1+e^{-(y_{it}-\alpha_i)/\beta_i}} \tag{3.4}$$

It won't make any difference once we look at the combined effect of simulation as well as conversion. As long as we simulate with the same μ and σ that we convert with, it doesn't matter what μ and σ are as long as σ is non-zero. We could use the values estimated from data, but we can also just use $\mu = 0$ and $\sigma = 1$.

3.2.1 Proof

Suppose that we started with price data, then converted to differences of logs and estimated a logistic on that data. Suppose we also estimated the mean and standard deviation, and those had values μ and σ. We have the y_{it},t=2,..,T, and we solve for x_{it}, t=2,..,T.

$$\int_{-\infty}^{x_{it}} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu)^2/\sigma^2} = \frac{1}{1+e^{-(y_{it}-\alpha_i)/\beta_i}} \tag{3.5}$$

We then estimate our variance-covariance matrix W. However, for now lets pretend we have just one series i. Now we come to the simulation point. And imagine that we simulate u_{it} using a normal distribution with $\mu = 0$ and $\sigma = 1$.

Now we could calculate

$$x_{it} = \mu + \sigma u_{it} \tag{3.6}$$

We now calculate the cumulative distibution function, F using

$$F = \int_{-\infty}^{x_{it}} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu)^2/\sigma^2} \tag{3.7}$$

We can substitute for x_{it} as

$$F = \int_{-\infty}^{\mu+\sigma u_{it}} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu)^2/\sigma^2} \tag{3.8}$$

Now to calculate this integral, we tranform z to a standard normal. We do this by setting

$$v = (z - \mu)/\sigma \tag{3.9}$$

Making this transformation, we find that

$$F = \int_{-\infty}^{u_{it}} dz \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}z^2} \tag{3.10}$$

Now this is the same thing as if we had just simulated u_{it} and then used this cumulative distribution to calculate F.

The value of y_{it} depends on F, α_i and β_i not on how F was calculated, as long as it produces the same F. So if we simulate u as mean 0 variance 1, and then transform u to some x which has mean μ and σ and calculate F using the same μ and σ its the same as if we calculate F using u and using $\mu = 0$ and $\sigma = 1$.

3.2.2 Extension to other distributions

The proof just given that the conversion from the normal to the non-normal distribution was independent of the normal's mean and standard deviation as long as the same parameters of mean and standard deviation are used to generate the normal deviates as to do the transformation did not rely on the characteristics of the logistic, and the same proof applies to any non-normal distribution.

3.3 Gamma Distribution

Reference Morris DeGroot Probability and Statistics p236. The probability density for the Gamma Distribution is

$$g(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{3.11}$$

for $x > 0$ and 0 otherwise. Here $\Gamma(\alpha)$ is the Gamma Function, which generalizes the factorial, $\Gamma(n) = (n-1)!$.

This gives us a 2 parameter distribution. We can compute the cumulative distribution function as

$$G(y|\alpha, \beta) = \int_0^y dx f(x|\alpha, \beta) \tag{3.12}$$

So that

$$G(y|\alpha, \beta) = \int_0^y dx \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{3.13}$$

We then convert from normals to this Gamma by first simlating a normal deviate u with cumlative density

$$F(u;\mu,\sigma^2) = \int_{-\infty}^{u} dz \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{1}{2}(z-\mu_i)^2/\sigma^2)} \tag{3.14}$$

and then converting to y such that

$$G(y|\alpha,\beta) = F(u;\mu,\sigma^2) \tag{3.15}$$

We have to interpret the results somewhat differently though. Since y is between 0 and ∞ it can be interpreted as the new price divided by the old price. In the case of the Logistic Distribution, we interpreted y as the change in the log of the stock price. Now we have to interpret it as the new stock price divided by the old stock price.

3.4 Beta Distribution

The beta distribution has probability density function

$$g(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{3.16}$$

for $1 > x > 0$ and 0 otherwise.

To represent stock returns with this distribution it is necessary to convert the stock return or stock price to a variable between 0 and 1 and vice versa. We can do this by taking a lognormal variable and converting it into a unit deviate by calculating u from x by

$$u = F(x;\mu^T,\sigma^T) \tag{3.17}$$

or given x, solving for u. The superscript T reminds us these are the transformation parameters. One choice is $\mu^T = 0$ and $\sigma^T = 1$.

Here F is the cumulative distribution of some lognormal variable. This can differ from the normal used for the generation of the random normals.

If we generated a normal x, converted that to u and then from that u to y for the Beta and then from that y back to x' as a normal, then we would have that x' was a normal determined essentially by the initial x directly. So we need a different method to generate a distribution of data.

3.4.1 Alternative generation of multivariate vector

We can instead use the univariate distributions of logistic, beta, gamma, etc and generate uniform deviates u_i. We can then transform these to normals x_i. We can then take a linear combination of these x_i, $y_j = \sum_i R_{ji} x_i$, where R is some matrix. Note the dimension of the x vector can be larger than the dimension of the sub-vector of y we wish to use. We can now treat the y_j as changes in logs of stock returns. Or we could transform them to be between 0 and ∞ and interpret those as ratios of prices.

In this approach we have a fundamentally discrete time generation process.

Chapter 4

Regime Switching

4.1 2 Regime Switching Model Lognornmal

For this model, let the parameters be μ_i and σ_i in state i, i=1,2. The parameters of the model shift at random times. Let ρ_{ij} be the transition rate per unit of time of going from state j to state i.

The lognormal model is

$$dx = \mu_i dt + \sigma_i dz \tag{4.1}$$

when in state i.

4.2 Estimating the parameters

The parameters are estimated using maximum likelihood as in Hamilton, Time Series Analysis, Chapter 22. See for example Mary Hardy's North Ameican Actuarial Journal paper for an application to the two state regime switching lognormal.

4.3 Estimating initial state

The initial state can be represented as known in one of the regimes, or as a probability density over them.

4.3.1 maximum likelihood

Can use the maximum likelihood method in Hamilton. This produces the probability density.

4.3.2 Trailing Data

We can also use trailing data, especially daily data if available. We estimate the parameters in the previous N days, and then use a simple comparison to the two states of the model and use the one that is closest. Alternatively, a weight can be given to each of the possible states, e.g. if the trailing vol is 20 percent and there are two vol states of 15 and 25 percent, one could use a weighted average, computed in standard deviation or variance for example as the measure.

4.4 Regime Switching Non-normal Models

We can generalize the discussion of the last section to cover the other univariate models besides the lognormal. Let the model have a parameter vector η, with elements η_i, i=1,..,n_η, where n_η are the number of elements in the vector η. Let there be n_r regimes of the parameters or discrete parameter sets.

Using the same methods as for the lognormal model other models can be estimated. This is done using maximum likelihood as in Hamilton.

4.5 Advantages of Regime Switching

4.5.1 Data description

There Regime Switching Model (RSM) describes the effect in data of periods of different values of parameters. Periods of high volatility in particular are important in modeling indices, especial equity indices. The Regime Switching Model does more than model the fat tails of the data, it also models the tendency for one fat tail event to be followed by another, i.e. a switch in regime in which large losses or gains are more likely. This is important to option writers because such gains or losses have a non-linear effect on the option price. This non-linear effect and the probability of greater price changes then results in diffrent prices for the options, so that their expected return in the risk neutral probability is the same. In particular,

the optionality of the option, its convexity is worth more when volatility is higher. This leads to a higher price of the option, put or call.

4.5.2 Risks to Option Writers Modeled by Regime Switching

Option writers, such as by variable annuity sellers, have the risk that volatility will change from what it is or they expect. When this happens, the Greek's of the options written change. This means that after a change in parameter values, the change in option prices for a given change in the stock price will be different. As a consequence, hedges set up to match Greeks, i.e. to match risk of the asset portfolio to the liability portfolio will no longer match. This is because the asset portfolio was chosen for a given set of Greeks. When the parameters change the asset and liability portfolio options all change in their Greeks, but because the Greeks are non-linear functions of the parameters and the mix of options in the asset and liability portfolio are different the options will change their prices from the parameter change and their Greeks, so that there is a discrete change in price from the parameter change and a discrete change in Greeks. This can cause a loss itself and also means the asset portfolio no longer hedges the liability portfolio.

The changes in prices and Greeks from the parameter change is a risk that can be analyzed using the regime switching model. Moreover, the change in hedge portfolio after the jump to one more appropriate to the new regime of parameters can be analyzed using simulations based on the regime switching model. One can estimate the cost for example of changing the hedges after a parameter change. Moreover, one can estimate the cost of the parameter changes and develop a portfolio to partially hedge this.

4.6 Canadian and US Insurance Regulation

Canadian and draft US C3 Phase II regulations for capital have a simulation option or requirement (use of tables is allowed in Canada). This does not require use of regime switching, although the calibration table in Canada was developed using regime switching analysis. Distributions with fat tails meet the calibration table with parameters that don't have super fat tails. By using a high enough volatility the lognormal will meet the calibration table. The calibration table requires the probability for large returns or large losses of specified size to exceed certain minimum probabilities. If one uses a lognormal and increases its volatility to fit the table, to meet all the points one needs a very large volatility, so that one exceeds the required probability for several of the points in the table. With the regime switching or other fat tailed distributions one can come closer to just matching each point in

the calibration table.

Chapter 5

ESG

5.1 DMRP

Let u and θ follow the process

$$du = \kappa_1(\theta - u)dt + \sigma_1 dz_1 d\theta = \kappa_2(\theta_2 - \theta)dt + \sigma_2 dz_2 \tag{5.1}$$

under realistic probability. The correlation between dz_1 and dz_2 is ρ. Let the risk neutral version be

$$du = \kappa_1'(\phi - u)dt + \sigma_1 dz_1 d\phi = \kappa_2'(\phi_2' - \phi)dt + \sigma_2 dz_2 \tag{5.2}$$

Where the variable ϕ is related to θ by a linear transformation. The correlation between dz_1 and dz_2 is still ρ and the values of σ_1 and σ_2 are the same in real as risk neutral as is required from no-arbitrage.

The instantaneous short term interest rate is

$$r = e^u \tag{5.3}$$

The prices of bonds are solved from

$$\frac{1}{2}\sigma_1^2 B_{uu} + B_u \kappa_1'(\phi - u) + \frac{1}{2}\sigma_2^2 B_{\phi\phi} + B_{u\phi}\rho\sigma_1\sigma_2 + B_\phi \kappa_2'(\phi - u) - e^u B + B_t = 0 \tag{5.4}$$

The parameters of the real and risk neutral processes are estimated from historical data over different time periods starting in the 1950's to date for U.S. Parameters for other countries and for the Euro are also available as well as dual linked currencies such as US Yen or Euro Yen in which the exchange rate is part of the state vector v and multiple sets of u, θ or u, ϕ are used for each economy.

In simulating real probability scenarios we can simulate u and ϕ but with realistic probability parameters for this process, yet a 3rd process from the above. Given u and ϕ we then calculate bond prices and yields. Alternatively, we can simulate u and θ and then calculate ϕ and then calculate bond prices.

5.2 ESG^{TM} Simulation

We first consider the case without regime switching. Let the vector v be m by 1, and let its elements be $v_0 = u$, $v_1 = \phi$, let $v_{i+1} = x_i$, for i = 1,..,n where n is the dimension of the x vector, and we index x starting from 1. We now form the process

$$dv = (b + Av)dt + Gdw \tag{5.5}$$

where b is m by 1, A is m by m and G is m by m. The vector dw is m by 1 and is a vector of mean 0, variance dt independent Wiener processes. We simulate v over some set of time intervals, which we suppose for simplicity are of equal size, Δt. We assume we start at $t = 0$, and generate a time series v_t, t=0,..,T. So we simulate $v_1, ..., v_T$ and we start with v_0. For stock return elements, $i > 1$ $v_{i0} = 0$ and for the interets rate model we use $v_{00} = u$ and $v_{01} = \theta$. Here the first index of v is the time index t, and the second index the index j for the vector component.

The first two elements of b and the first 2 by 2 sub-matrix of A are formed so as to replicate the DMRP process above. All other elements of A and b are zero.

The correlation matrix GG' is estimated for the joint correlation matrix of u, θ and x, i.e. v.

We now simulate v and for each element of x, we have $x_i = v_{i-1}$ suppressing the t index, and remembering we are indexing x starting with 1 and v starting with 0.

We now convert each x_i to y_i using the univariate distribution for y_i. This can be normal or logistic or some other univariate distribution. We use the appropriate parameter vector b_i to do this, taking care our parameters are approriate for the time interval Δ. As proven earlier it doesn't matter what the mean and variance of the x_i are at each t, as long as we use the same ones for the conversion as for the simulation. We can in fact just use unit normals used in generating the w[t] vector.

5.3 ESG with Regime Switching

We have the vector of continuous variables

$$dv = (b + Av)dt + Gdw \tag{5.6}$$

In addition to this is the state index s, which we interpret as a set of a finite number of states. We have a mapping from s to b,A and G, assigning for each s, b(s), A(s) and G(s). Different values of s can then correspond to the different paramters of the univariate models for each index. We have a process on s, of the form $\rho(s', s)dt$ where $\rho(s', s)$ is the transition rate per unit time to go from state s to state s'. In addition the transformation functions from v to y are indexed by s in addition to their other parameters so $y = h(v, s, t)$ is the transformation function from v to y contingent on state s.

5.4 ESG Estimation

To estimate the correlation matrix of the ESG, we proceed along the lines of Chapter 2, but now include past values of u and θ estimated from the yield curve on past dates to get these values and then estimate the correlation matrix of u, θ and x.

As before, we estimate each x time-series, e.g. S & P 500 or Ford stock price first as a univariate distribution, e.g. Logistic and then use the estimated parameters to transform to the normals using the cumulative distribution method.

Appendix A

Notation

A.1 Basic parameters and variables

A.1.1 n

The number of individual data series, e.g. S & P 500, Ford, etc. These are numbered 1 to n.

A.1.2 T

The number of time periods, numbered from 1 to T.

A.1.3 x

An n by 1 vector from a multivariate normal distribution.

A.1.4 y

An n by 1 vector of variables transformed from x. These are the observables.

A.1.5 Σ

Correlation matrix that is n by n. Used to do simulation, and is the theoretical unobserved true correlation matrix of the x vector.

A.1.6 W

This is the estimated correlation matrix. It is used to do the actual simulation as if it was the theoretical correlation matrix when the software system is used in practice.

A.2 Returns

A.2.1 Log return

Given any time series, V_t, t=1,...,T, such as a price, or other variable which is non-negative, we can define its log return as u_t

$$u_t = lnv_t - lnv_{t-1} \tag{A.1}$$

for t=2,..,T. We thus "lose" a data point at t=1.

So for example, y_{it} might be the change in the natural logarithm of the change of Ford's stock price. We calculate it from the observed stock prices, P_t and P_{t-1} as

$$y_{it} = lnP_t - lnP_{t-1} \tag{A.2}$$

where i corresponds to Ford, say i=1 was Ford.

To go the other way, to get P_t given y_{it} we calculate

$$P_t = P_{t-1}e^{y_{it}} \tag{A.3}$$

We have to know P_{t-1}. If we are simulating we have to know P_0 say and then we can simulate for later t using $y_{i1},\ldots,y_{iT}$.